A rating scale for wildness and ease of handling laboratory mice: results for 21 inbred strains tested in two laboratories

By: Douglas Wahlsten, P. Metten and J. C. Crabbe

Wahlsten D, Metten P, Crabbe JC. (2003). A rating scale for wildness and ease of handling laboratory mice: results for 21 inbred strains tested in two laboratories. *Genes, Brain, and Behavior*, 2: 71-79.

Abstract:
Rating scales for difficulty in capturing and holding mice were devised that proved to be easy to use and highly sensitive to differences among mouse strains on the A and B priority lists of the Mouse Phenome Project. The simplicity of the scales makes it feasible to rate wildness during behavioral test sessions without adding much to testing time or distracting the technician from the principal task at hand. Overall wildness and placidity ratings obtained by combining capture and hold ratings provide a good impression of the difficulties encountered while working with lab mice in the course of complex experiments. Ratings of 21 inbred strains during the course of 15 behavioral tests in two laboratories demonstrated that the SPRET/Ei, PERA/Ei, CAST/Ei and SWR/J strains were particularly difficult to handle. The NOD/LtJ strain posed no special challenge in the Edmonton laboratory but was very difficult to handle in the Portland lab. The rating scales should be useful for judging the difficulties in working with novel targeted or induced mutations in mice as well as effects of a variety of environmental treatments or drugs.
Keywords: Emotionality, gene-environment interaction, laboratory environment, Mouse Phenome Project, vocalization, wild-derived inbred strain

Article:
Reaction of laboratory rodents to human handling is a behavioral phenotype of special interest because the Mouse Phenome Project (MPP, Paigen & Eppig 2000; see http://www.jax.org/phenome) has brought researchers into close contact with several strains of mice that are recently wild-derived from species other than Mus musculus domesticus. Although some behavioral studies of mice captured in the wild (Smith & Connor 1974) as well as the wild-derived inbred strains CAST/Ei (Koide et al. 2000; LeRoy et al. 1998) and SPRET/Ei (Laghmouch et al. 1997) made no mention of difficulties working with these animals, our experiences and tales from many colleagues suggest otherwise. Mouse tales, despite their charm, are subjective and difficult to grasp. We need more objective criteria to aid comparison between different laboratories or even different technicians in the same lab. In this paper, we present a rating scale for wildness of mice and apply it to 21 inbred strains, including four that were wild-derived, tested simultaneously in our two laboratories.

Our rating scale for wildness or ease of handling was devised with three applications in mind. Firstly, wildness is a phenotype worthy of genetic investigation in its own right (Stone 1932) and as a component of the domestication process (Richter 1954). Secondly, results of common behavioral tests that involve extensive handling of the subjects by an experimenter may be altered or difficult to interpret when the mouse reacts strongly to handling. Finally, extreme reactions to handling may warrant exclusion from further testing. Presuming that most researchers are not interested in wildness per se but rather in its tendency to impair validity of other tests, we also tried to make the scale relatively simple and easy to use.

Although several rating scales for wildness have been described, we found that these could not be adapted for several reasons. Some involved bizarre attempts to provoke responses, such as when the experimenter made ‘a

clicking sound with his tongue and teeth'(Coburn 1922), tickled the animal's nose, or blew 'two puffs of strong cigar smoke' into the cage (Keeler 1942). Early scales rated wildness and savageness or even timidity separately (Coburn 1922; Stone 1932; Yerkes 1913), but their distinctiveness as psychological constructs is difficult to defend. Several scales were originally devised to compare tame, laboratory strains to animals recently captured in the wild state (Coburn 1922; Galef 1970; Stone 1932; Yerkes 1913). Consequently, the upper end of the scale often involved extreme reactions that we have never seen in any lab mouse (Coburn 1922), while the lower end of the scale generally failed to discriminate among more subtle differences in behavior. For example, Coburn's (1922) 6-point scales of wildness and savageness assigned scores of 0 to all 'tame' albino mice.

Table 1: Rating scales for mouse wildness and ease of handling. (a) Capture – from the home cage, the testing cage, or a test apparatus

Score	Frequency	Behavior	Comments
0	4766	Minimal resistance to capture	
1	542	Evades touch by running around cage	Must complete at least one circuit of cage, going end to end twice or following along the full perimeter
2	191	Jumps onto wall of cage	While being pursued; does not apply to cage mates awaiting their turn; mouse is often captured by the tail on the wall
3	60	Jumps completely out of cage, onto table	Either large jump from cage floor or short hop down from wall; often is captured after landing on table
4	36	Runs from vicinity of cage	With or without experimenter in hot pursuit
5	48	Jumps off table or apparatus onto floor	Usually while being pursued; could happen with one large jump from the cage floor; often captured soon after landing
6	107	Runs around room	Need not complete a full circuit; anything requiring active pursuit to recapture

(b) Holding – when being picked up and placed elsewhere or held for injection

Score	Frequency	Behavior	Comments
0	4527	Minor or no struggle	
1	777	Squeaks or squeals	Must be clearly audible
2	341	Vigorous struggle or twisting/shaking	May attempt to pull away or climb up tail; can occur without any squeak being heard
3	51	Attempts to bite	Jaws open near hand or forceps
4	52	Bites experimenter	Teeth make contact with glove or skin or forceps; need not penetrate or draw blood

Note: Frequencies based on observations of 392 mice from 21 inbred strains on 15 tests made in both Edmonton and Portland. The total number of capture events (5750) was less than $392 \times 15 = 5880$ because not all 15 tests were completed for some mice. No trial was terminated because of excess wildness. The number of holding events is two less than 5750 because two mice were never re-captured on one test. One mouse spent the night inside a door that was found to have a small hole in its bottom edge. None of the 52 bites actually penetrated the plastic glove and drew blood from the experimenter.

Several of the scales were devised for wild rats and therefore included threat behaviors that are extremely rare in mice, such as gnashing of the teeth, hissing and lunging at the experimenter (Stone 1932; Yerkes 1913). Another lacked sufficient precision in the definitions of behavior (Manser et al. 1994). Several scales included defecation and urination (Galef 1970; Stone 1932), emotional responses that are frequent in any novel or stressful environment and do not create special difficulties for the experimenter.

The rating scales

Rather than choosing test items to measure some underlying psychological factor or construct, we designed the scales to follow closely the operations typically performed during a test. Thus, there are two scales, one for capturing the animal and the other for holding it. Many years of experience have taught us that some mice are relatively easy to capture but protest loudly and struggle valiantly when picked up and held for a few seconds or injected with a hypodermic needle. Likewise, some flee at the slightest touch of the tail or even sight of the hovering hand of the experimenter but then submit once firmly grasped. Within each scale, ratings were ranked in order of the severity of the animal's reaction (Table 1). For the capture scale, one number was assigned, representing the worst offence against orderly conduct committed on a trial. The score assigned to an animal on one trial was associated with the time required to effect a capture, but we found that keeping track of capture

time took too much technician time and was a distraction for the task at hand. Similarly for the holding scale, the worst behavior was scored. Time in seconds was not pertinent to the holding scale.

Because of the intimate relationship between human and mouse during handling, the scales rate both the subject and its master, and data will be most readily interpreted when the technician is highly skilled after months or years of working with lab mice. In a letter to Coburn (Coburn 1922), the legendary Abbie Lathrop, the supplier of ancestral breeding stock for several important mouse strains, observed: 'Every one who comes here and tries to handle mice gets bitten even by tame mice, because of grasping the tail too tightly'. At the same time, if mice of an inbred strain were randomly assigned to a technician in a single lab, the scales could be used to rate skill at handling mice. An experienced technician will of course anticipate that certain wild-derived strains such as SPRET/Ei will be difficult to capture and hold, and a certain degree of compensatory effort may be expected. There is a limit to the quickness of movements that are useful in capturing mice, however. For SPRET/Ei, there is no more dependable way to provoke an attempt to flee than making a fast movement with the hand towards the mouse.

The ratings of capture and hold can be made for a single occasion, such as giving an injection of a drug. They can also be applied to a session involving several training trials on the same task. Table2 indicates the sequence of tests that were done over a 16-day period in our behavioral studies for the MPP. For example, in one water escape trial the mouse is captured in its holding cage, held by the tail and transferred to the water tank, then picked up from the escape platform and returned to the holding cage. Thus, four trials involve eight episodes of capture/hold. Each episode could be rated separately and then summed or averaged, but we found this to be too distracting and labor intensive. Instead, the technician assigned a score to a mouse that represented the worst conduct on any of the eight episodes. This ploy was essential on a complex test such as the accelerating rotarod where 10 trials were done with four mice at the same time. All the technician needed to do was jot down any animal's non-zero rating on the first trial and then revise it only if the rating was worse on a subsequent trial. If no non-zero score was written, the technician thus recorded zero for capture and handling of that mouse. Comparisons of wildness ratings for different kinds of tasks are therefore meaningful only when they entail the same number of capture/hold episodes. To use our rating scales, researchers should adhere closely to the definitions given in Table 1, but there is great flexibility in how ratings are combined for multiple episodes of handling.

Materials and methods

Mice

The inbred strains listed in Table 3 were obtained from the Jackson Laboratory, Bar Harbor, Maine, at 7weeks of age, tested for several behaviors at 10–1 2weeks of age, and then examined for brain defects at 12.5weeks. The mice were sent in five simultaneous shipments to each lab, with each shipment usually providing two or three mice per strain per laboratory. For any one shipment to a lab, all mice of a given strain were the same sex and could have been littermates. If males of a strain were sent to one lab in a shipment, females were usually sent to the other lab. Thus, no more than four or five mice per strain could have come from the same litter, and the total sample of about 10 males and 10 females must have been taken from at least five litters. The complement of strains in a single shipment was not identical for the two labs, but the net result of the five shipments was nearly equal sample sizes in the two labs. When enough mice of the appropriate age were not available at Jackson Laboratories, more than two or three would be sent in a later shipment for that strain. The actual numbers in each shipment ranged from 31–44 for Portland and 27–49 for Edmonton. Mice were present in at least four shipments for 15 of the 21 strains sent to Portland and 18 strains sent to Edmonton. The greatest deviation from the intended design was for the SPRET/Ei strain where only one shipment of three females were sent to Portland and two shipments (two males, three females) were sent to Edmonton. Mice were missing from three shipments for the MOLF/Ei strain sent to Portland and BTBR to Edmonton.

Conditions in the colony rooms were similar at the two sites. Animals were housed two or three per cage in plastic shoebox cages with 2.5 mm Bed-o-cob bedding (Andersons, Maumee, OH, USA). They had free access to 5001 lab chow (PMI Nutrition International, Brentwood, MO, USA) and local tap water. Colony room lights

went on at 06.00 h and off at 18.00h. Certain wild-derived strains (MOLF/Ei, PERA/Ei, SPRET/Ei) were maintained with filter tops on the cages in Edmonton because they came from colony rooms at the Jackson Laboratories that housed mice known to carry Pasturellapneumotropica. All cages hadfiltertops in Portland.

Table 2: Tests administered to mice and rated for wildness

Test #	Day	# of handlings	Kind of test
1	Monday	2	One 5 min trial of open field activity
2	Tuesday	2	One 5 min trial on elevated plus maze
3	Wednesday	11	Ten trials on rotarod acquisition
4	Thursday	1	Saline injection (intraperitoneal)
5	Thursday	7	Six trials of rotarod testing (three before injection)
6	Friday	1	Ethanol injection (intraperitoneal)
7	Friday	7	Six trials of rotarod testing (three before injection)
8	Monday	6	Three trials of water escape pretraining
9	Tuesday	8	Four trials of water escape training
10	Wednesday	8	Four trials of water escape training
11	Thursday	8	Four trials of water escape training
12	Friday	8	Four trials of water escape training
13	Friday	2	Weighing on triple-beam balance
14	Monday	2	Weighing after free feeding over weekend
15	Tuesday	2	Weighing after 24 h food deprivation

Table 3: Strains means ± standard deviation for measures of wildness

Strain	N	Wildness	# placid tests in 15	Capture	Holding	# tests in 15 with squeak
129S1/SvImJ	20	0.16±0.11	11.8±2.7	0.01±0.03	0.16±0.12	2.8±2.1
A/J	18	0.08±0.12	13.5±1.8	0.02±0.09	0.05±0.05	0.7±0.7
AKR/J	17	0.78±0.31	4.3±2.5	0.05±0.16	0.72±0.19	**9.6±2.4**
BALB/cByJ	20	0.13±0.13	13.0±1.6	0.04±0.05	0.09±0.11	1.2±1.3
BTBR T+tf/tf	18	0.48±0.34	11.1±2.4	0.23±0.27	0.25±0.14	2.5±2.4
C3H/HeJ	18	0.36±0.22	10.5±2.7	0.18±0.25	0.19±0.12	1.7±1.3
C57BL/6J	18	0.27±0.20	12.0±1.7	0.08±0.13	0.20±0.12	1.3±0.8
C57L/J	18	0.21±0.10	12.1±2.0	0.05±0.10	0.16±0.07	1.8±0.7
C58/J	19	0.30±0.25	10.7±2.9	0.18±0.17	0.13±0.10	0.7±0.9
CAST/Ei	19	**1.35±0.78**	5.4±3.7	**0.97±0.74**	0.38±0.42	0.4±0.7
DBA/2J	18	0.66±0.63	8.7±3.9	0.28±0.43	0.36±0.28	**4.6±2.8**
FVB/NJ	21	0.61±0.45	9.0±3.2	0.37±0.31	0.23±0.18	2.1±2.1
MOLF/Ei	14	0.60±0.54	12.1±2.4	0.30±0.43	0.30±0.22	0.0±0.0
NOD/LtJ	24	**1.23±1.26**	7.1±4.9	0.57±0.72	0.65±0.65	1.5±1.6
NZB/B1NJ	18	0.14±0.13	14.0±0.7	0.09±0.14	0.05±0.06	0.3±0.5
PERA/Ei	22	**2.11±1.19**	6.1±3.8	**1.53±1.07**	0.60±0.29	1.5±1.3
PL/J	20	0.05±0.11	14.0±1.9	0.04±0.11	0.00+0.02	0.0±0.2
SJL/J	20	0.23±0.21	11.2±3.4	0.03±0.06	0.20±0.17	2.3±2.1
SM/J	24	0.15±0.19	13.2±1.9	0.09±0.15	0.05±0.09	0.2±0.4
SPRET/Ei	7	**3.74±0.71**	4.3±3.5	**2.68±0.91**	1.06±0.27	0.3±0.5
SWR/J	20	**1.38±0.61**	5.5±2.72	**0.62±0.43**	0.76±0.32	**5.6±2.8**

Note: Values averaged over 15 tests listed in Table 2. Sample sizes for certain measures of a few strains were slightly less than the given value because one or more occasions in the 15 tests yielded missing data. Bold figures denote particularly high values.

Behavioral testing

During behavioral testing, mice were picked up by the tail by a technician wearing disposable plastic gloves; forceps were never used. The mouse in a holding cage was brought close to the apparatus before the animal was picked up, and the mouse was never carried by the tail for more than a few cm. Behavioral testing was done by two technicians in Portland and three in Edmonton, but technicians were not explicitly included in the design or balanced across tests and con-itions. In Portland, all trials of one kind of test (e.g. elevated plus maze) were administered by one technician and all trials of rotarod acquisition were given by another technician. Consequently, formal tests of variation due to technician skill were not feasible in this study. Each mouse was given the same sequence of behavioral tests listed in Table 2. Within a day, the order of testing mice of most strains and sexes was randomized, and the test order used on the first Monday was employed on all subsequent

days. The wild-derived strains were tested at the end of each day, owing to concerns about completing testing in time in the event of mice escaping and evading capture for a significant amount of time. All mice in one shipment were run through a specific test on the same day. Testing began at about 08.30 h and concluded by 17.00 h, so that all tests were given during the light phase of the day-night cycle. Details of the tests, including apparatus and procedural protocols, are provided in two recent publications (Rustay et al. in press; Wahlsten et al. in press).

Briefly, the open field was a 40 × 40 × 30 cm high clear plastic box with a clean paper floor that was fresh for each mouse. The device was enclosed in a sound-attenuating cubicle with a TV camera directly above the floor so that the mouse could be tracked automatically (VideoScan system from AccuScan Instruments Inc., Columbus, OH). One trial of 5 min duration was given.

The elevated plus maze had black plastic arms 5 cm wide and 30cm long, and the device sat on a pedestal 50cm above the floor. The two closed arms had 15cm high clear plastic walls, whereas open arms had a 0.25-cm clear plastic lip. One 5 min trial was given. Animal movement was tracked with the VideoScan system.

Motor coordination was assessed with the AccuRotor device from AccuScan that was modified to have a fall height of 63 cm and rod diameter of 6.5 cm. The rod surface was covered with 320 emery paper affixed with rubber cement. Acceleration rate was 20 r.p.m./min. Acquisition consisted of 10 trials with a 30 second intertrial interval (spent in the bedding trough). The next day, the mouse received three further acquisition trials and was then injected with physiological saline, placed into a holding cage for 30 min, and given another three trials. The day after that, there were three trials followed by an injection of 2.0 g/kg ethanol, a 30 min delay, and then three more trials.

Water maze testing was done in a 70cm diameter tank filled with 26ºC water. An escape platform of 10cm diameter was placed 0.5 cm beneath the surface of water rendered opaque with white and blue tempera paint. The maze had four arms 10.5cm wide made from clear plastic. Three pretraining trials were given on Monday, and then four training trials were given each subsequent day with a 30 second intertrial interval and 60 second trial limit. On each trial, the mouse was placed at the center of the tank facing one of three random compass positions (never the correct direction), and it spent the intertrial interval in a holding cage.

Results

The average strain ratings on the capture and hold scales are shown in Fig. 1 for each of the 15 test occasions described in Table 2. Several things are obvious from the graph. Firstly, several strains were consistently easy to capture and hold (A/J, BALB/cByJ, NZB/B1 NJ, PL/J and SM/J). Secondly, several strains were always easy to capture and resisted being held only during an intraperitoneal injection (129S1/SvImJ, C3H/ HeJ, C57BL/6J, C57L/J, DBA/2J, MOLF/Ei and SJL/J). For strains where average holding scores exceeded 1.0, resistance was more than simply squeaking. Thirdly, wild-derived strains CAST/Ei, PERA/Ei and SPRET/Ei as well as laboratory strains FVB/NJ, NOD/LtJ and SWR/J were especially challenging. The difficulties working with SWR/J were similar to those reported 30years earlier (Wahlsten 1973). Fourthly, most strains that were initially difficult to capture showed a gradual habituation to handling over several test sessions, with the exception of strains NOD/LtJ and SWR/J. And finally, as training progressed, the strains CAST/Ei, PERA/Ei and SPRET/Ei became quite easy to hold but remained very difficult to capture. Mice of these strains commonly made the first move to evade capture when the experimenter approached, before the animal's tail was touched. Especially in the first week of testing, mice of these strains added substantial amounts of time to the testing day because of their evasiveness.

A composite wildness score may be obtained by adding the capture and hold scores, yielding a value ranging from 0 (placid, no resistance of handling) to 10 (runs around the room, then bites when finally captured). The average wildness rating is shown in Table 3, averaged over the 15 test occasions listed in Table 2. Placidity, the number of tests out of 15 when the wildness score was 0, was inversely related to wildness. These two values provide a good impression of the degree of difficulty in handling the various strains that may be expected in an

extensive test battery. Average capture and hold scores along with counts of squeaking reveal that strain AKR/J had a relatively high wildness score primarily because it squeaked when held. Wild-derived mice, on the other hand, generally squeaked very little but were difficult to capture (except MOLF/Ei).

There was considerable individual variation in wildness from one testing day to the next. Nevertheless, the reliability of the rating scales was reasonably good. For all 344 mice having complete data, the Pearson correlation of capture scores was r = 0.56 on days 1 (open field) and 2 (plus maze), r = 0.68 for the rotarod acquisition day (10 trials) vs. rotarod saline day (6 trials) and r = 0.55–0.74 for the four days of water escape training. Correlations for holding scores were r = 0.68 for day 1 vs. day 2, 0.68 for rotarod acquisition vs. saline and r = 0.42–0.58 for the four days of water escape training. The correlation of wildness scores for the three days of rotarod testing and the four days of water maze training was r = 0.63. For the total sample, all correlations greater than 0.25 were significant at P < 0.01.

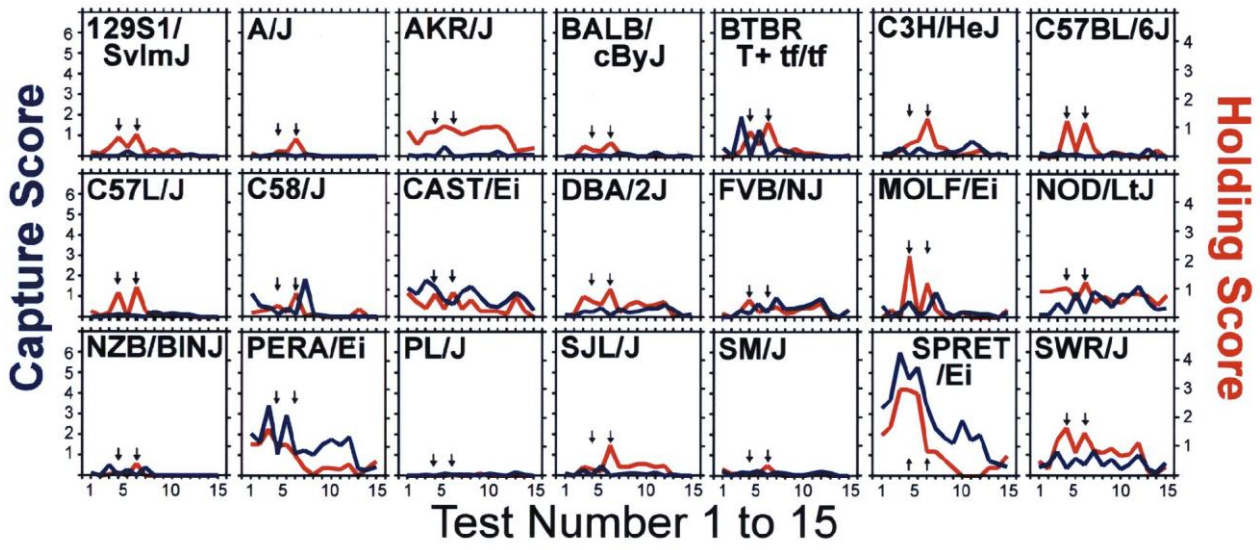

Figure 1: Average capture (blue) and holding (red) ratings during 15 behavioral tests (Table 2) for 21 inbred mouse strains. The two small arrows indicate conduct during intraperitoneal injections of saline on Thursday (test 4) and ethanol on Friday (test 6).

The rating correlations among the 21 strain means, each based on about 20 mice (Table 3), were generally very high. The Pearson correlation of capture scores was r = 0.96 on days 1 (open field) vs. 2 (plus maze), r = 0.97 for the rotarod acquisition day (10 trials) vs. rotarod saline day (6 trials) and r = 0.88–0.96 for the four days of water escape training. Correlations for holding scores were r = 0.92 for day 1 vs. day 2, 0.95 for rotarod acquisition vs. saline and r = 0.88–0.94 for the four days of water escape training. The correlation of wildness scores during open field and plus maze testing vs. the three days of rotarod testing was r = 0.95 and vs. the four days of water maze training was r = 0.79. As shown in Fig. 2, the most extreme strains were wild derived, but the strain correlations were quite high even for the 17 strains that were not wild derived, despite the narrower range of scores. With 17 strains, all correlations greater than 0.61 were significant at P < 0.01.

Within-strain correlations across different test occasions were of course considerably lower because of restricted range and non-significant for many small samples. For some strains the correlations could not be computed because at least one of the measures had no variance (all scores = 0). Individual mice varied considerably from trial to trial and were, thankfully, not consistently wild on every test. Correlations were generally higher for variables formed by combining data across several trials. For example, the median within-strain correlation between wildness scores on the rotarod and water escape tasks was r = 0.36 and this correlation was particularly high for the 4 wild-derived strains that tended to have substantial within-strain variation (Table 3).

Analysis of variance (ANOVA) was done for a complete strain × sex × site design without the SPRET/Ei strain for which only females were tested in Portland. Because these analyses revealed non-significant (P>0.01) or

very small effects of sex or interactions with sex, data were pooled over sex and ANOVAs were done for 21 strains × 2 sites. To reduce the mean-variance correlations across strains, wildness, capture and hold scores were transformed to the log10 scale. As indicated in Table4, the strain differences that are obvious in Fig. 1 were highly significant and very large indeed for all variables. Significant site effects were also detected, but the interactions of strain and site were especially interesting. As shown in Fig. 3, the site main effect for wildness and placidity was considerably less than the strain–site interaction because the NOD/LtJ mice in Portland were considerably more wild than in Edmonton, whereas Edmonton encountered greater difficulty than Portland in handling the PERA/Ei, SPRET/Ei and SJL/J strains. The lab difference for SPRET/Ei mice was based on a small sample size consisting of only one shipment to Portland and two shipments to Edmonton, and hence it cannot be readily interpreted. Among the four wild-derived strains, a marked lab difference was observed only for PERA/Ei.

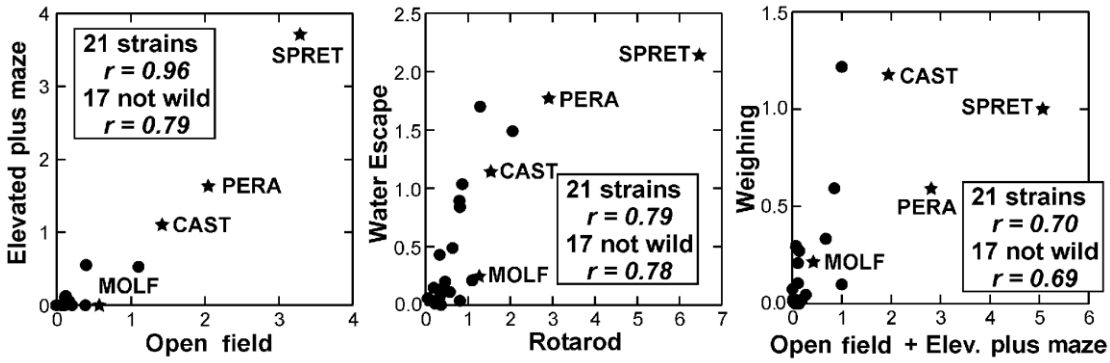

Figure 2: Scatterplots of strain mean wildness scores during the administration of different tests. Wild-derived strains are shown as stars. Strain correlations remained very high when the four wild-derived strains were omitted from the calculations, although they appear reduced because they are compressed into a narrower range when the wild-derived mice are included in the graph. Open field and elevated plus maze tests involved a single trial each, whereas rotarod acquisition involved 10 trials and water escape training involved 16 trials. Weighing was done three times.

The lab difference for the NOD/LtJ mice was particularly striking; they ranked second in wildness in Portland but only ninth in Edmonton. The two sites received equal numbers of NOD/LtJ mice in all five shipments, and in every shipment the wildness ratings were considerably higher in Portland, indicating this was a robust strain-specific effect of lab environment. Neither lab had previous experience with or expectations about NOD/LtJ regarding ease of handling, and the technicians handling the mice did not know the strain membership at the time of handling because they were albino and several albino strains were tested in random order the same day.

Table 4: Results of analysis of variance for measures of wildness and ease of handling

Measure	Total N	Strain (d.f. = 20)	Site (d.f. = 1)	Strain × Site (d.f. = 20)	Multiple R^2
Log (Wildness)	342	$F = 42.4$ $P < 0.000001$ est $\omega^2 = 0.72$	$F = 4.4$ $P = 0.04$ est $\omega^2 = 0.01$	$F = 5.7$ $P < 0.000001$ est $\omega^2 = 0.23$	0.766
Log (Capture)	344	$F = 31.4$ $P < 0.000001$ est $\omega^2 = 0.65$	$F = 19.0$ $P = 0.00002$ est $\omega^2 = 0.06$	$F = 5.2$ $P < 0.000001$ est $\omega^2 = 0.21$	0.727
Log (Hold)	345	$F = 35.0$ $P < 0.000001$ est $\omega^2 = 0.68$	$F = 10.4$ $P = 0.001$ est $\omega^2 = 0.03$	$F = 10.5$ $P < 0.000001$ est $\omega^2 = 0.37$	0.746
# squeaks	393	$F = 36.0$ $P < 0.000001$ est $\omega^2 = 0.65$	$F = 1.9$ $P > 0.05$ est $\omega^2 = 0.003$	$F = 1.6$ $P > 0.05$ est $\omega^2 = 0.03$	0.684
# placid tests	393	$F = 27.4$ $P < 0.000001$ est $\omega^2 = 0.59$	$F = 7.0$ $P = 0.009$ est $\omega^2 = 0.02$	$F = 5.3$ $P < 0.000001$ est $\omega^2 = 0.19$	0.657

Note: Data were combined for males and females because the sex difference and interactions with sex were almost always not significant ($P > 0.01$). est ω^2 estimates the proportion of variance attributable to the effect in question when only that effect is compared with variation within groups. Multiple R^2 is the proportion of total variance accounted for by the three effects in combination.

The large strain differences and strain by lab interactions were not simply the result of a few extreme-scoring wild-derived strains. The range of scores for the 17 strains that were not wild-derived appears to be rather narrow in Figs2 and 3, but if the wild-derived strains are omitted, it is clear that the other 17 strains ranged widely in wildness scores (see Fig.4). When the ANOVAs were done without the four wild-derived strains,

results were very similar to those shown in Table3 for all 21 strains; strain differences and interactions between strain and site (except for squeaking) were clearly significant (P<0.000001), and multiple R^2 values ranged from 0.53 to 0.75.

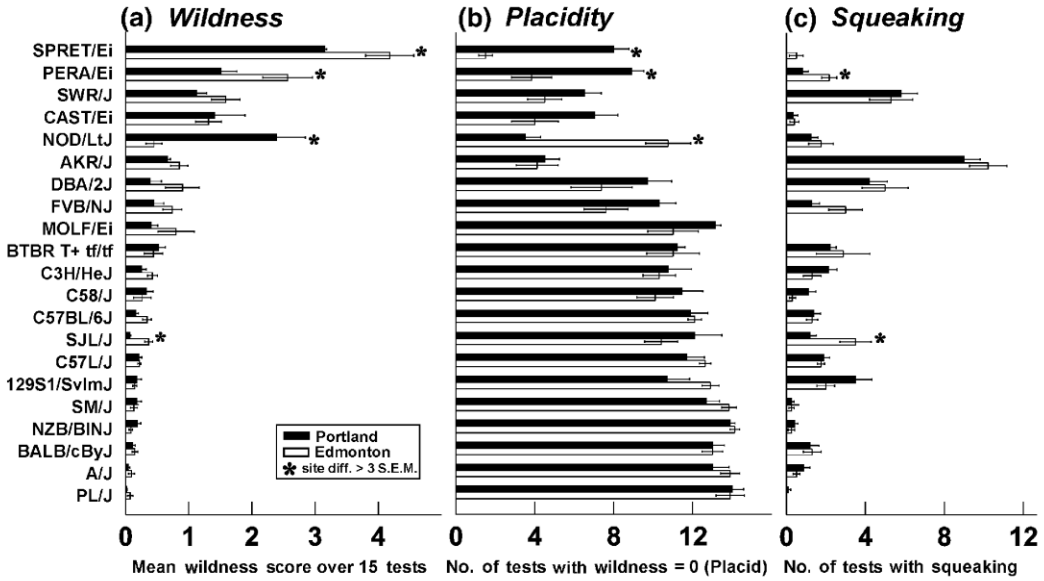

Figure 3: Strain means and standard error bars for 21 inbred mouse strains tested simultaneously in two laboratories. Strains are arranged in rank order by average wildness score. Wildness and placidity ratings were inversely proportional. Squeaking during handling was generally low in wild-derived strains that were otherwise high in wildness (SPRET/Ei, PERA/Ei, CAST/Ei). * indicates a strain where the means of the two labs differed by more than three standard errors. The large lab difference for the NOD/LtJ strain is especially noteworthy.

Discussion

The rating scales for difficulty in capturing and holding mice were easy to use and highly sensitive to differences among a wide variety of mouse strains. The scales revealed substantial variation among common laboratory strains that were not wild-derived and are not limited to detection of extreme wildness. The simplicity of the scales makes it feasible to rate wildness during behavioral test sessions without adding much to testing time or distracting the technician from the principal task at hand. Overall wildness and placidity ratings obtained by combining capture and hold ratings provide a good impression of the difficulty encountered in the course of complex behavioral experiments. Reliability of the scales on repeated tests was high for our large set of strains, although it will be somewhat lower in studies that do not employ the higher scoring wild-derived strains.

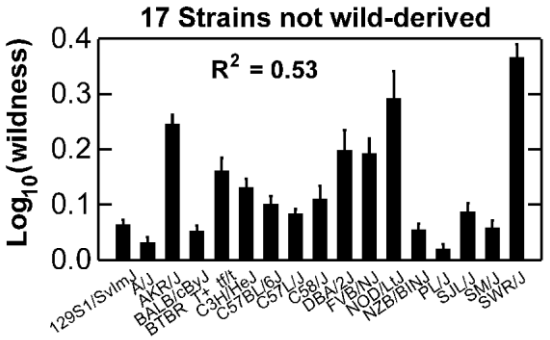

Figure 4: Mean log_{10} wildness score and standard error of the mean for 17 inbred strains that were not wild-derived. The statistical strength of the strain variation (multiple R2) was very large, and many strains differed by more than three standard errors of the mean.

This study was not designed to identify the underlying psychological processes that give rise to high levels of wildness, and it could not reveal whether a placid animal was unusually fearful or timid. Given the genetic complexity of differences between inbred strains, wildness is not likely to be a unitary trait, and two strains can probably have relatively low scores for different reasons. Inbred strains offer an especially attractive approach

to this question because strain correlations can be computed between wildness scores and a wide range of other behaviors assessed on the same strains as part of the Mouse Phenome Project.

The rating scales should be useful for judging the difficulties in working with novel targeted or induced mutations in mice. Our results for 21 inbred strains will become part of the Mouse Phenome Database and can serve as norms for judging new mutants, other strains, or even environmental treatment effects. Because we worked with two wild-derived strains (PERA/Ei, SPRET/Ei) that were definitely a challenge for the technicians, the scores of those strains may be considered as levels worthy of concern in other mice. It would then be a straightforward matter to determine whether housing one mouse per cage induces wildness to a degree seen in wild-derived inbred strains, for example.

It is possible that a scale could be altered to detect even finer differences among the least difficult laboratory strains. Nevertheless, our experience indicates that mice in the low-scoring strains generally showed very similar patterns of behavior in response to human handling. Perhaps sensitivity to individual differences could be increased by testing each animal more often, but the results would depend strongly on the kind of test. Mice tend to habituate to handling if they are always treated gently, whereas a few electric shocks might reveal latent differences in wildness.

Ratings were similar in the two labs for most strains, but Edmonton obtained slightly higher scores than Portland when averaged over all strains. The present study cannot determine whether the lab effect arose from differences between technicians or the pre-testing lab environments. Future research might evaluate these alternatives by swapping technicians at the two sites. A study of this nature would benefit from randomized assignment of mice to shipping destination on the same day or shipping of matched littermates to the two labs.

Previous studies have sometimes employed repeated handling of rats to achieve a certain degree of tameness before commencing behavioral testing on a variety of tasks (Barnett 1958; Hughes & Boice 1973). This might be feasible for inbred mice, but the number of handling trials required would differ substantially among strains and could create or confound strain differences on other tests. We recommend that all strains be handled the same amount by expert personnel and that handling mice be minimized. If this is done, reaction to handling will constitute a phenotype rather than a genotype-dependent treatment effect. Whether the difficulties handling certain mice contaminates results of other tests can then be assessed, as a first approximation, by correlations between wildness scores and other test results, especially the within-strain correlations. Within-strain correlations will necessarily be low when mice differ little in wildness scores, but we have found that this lack of variance only occurs for unusually low levels of wildness where none of the animals challenges the human handler.

References

Barnett, S.A. (1958) Laboratory methods for behaviour studies of wild rats. J Anim Technician Assoc 9, 6–14.
Coburn, C.A. (1922) Heredity of wildness and savageness in mice. Behav Monographs 4,1–71.
Galef, B.G. Jr (1970) Aggression and timidity: responses to novelty in feral Norway rats. J Comparative Physiol Psychol 70, 370–381. Hughes, C.W. & Boice, R. (1973) Domestication, sophistication, and avoidance in Norway rats. J Comparative Physiol Psychol 84, 408–413.
Keeler, C.E. (1942) The association of the black (non-agouti) gene with behavior in the Norway rat. J Heredity33, 371–384. Koide, T., Moriwaki, K., Ikeda, K., Niki, H. & Shiroishi, T. (2000) Multi-phenotype behavioral characterization of inbred strains derived from wild stocks of Mus musculus. Mammalian Genome 11, 664–670.
Laghmouch, A., Bertholet, J. & Crusio, W. (1997) Hippocampal morphology and open-field behavior in Mus musculus domesticus and Mus spretus inbred mice. Behav Genet 27, 67–73.
LeRoy, I., Roubertoux, P.L., Jamot, L., Maarouf, F., Tordjman, S., Mortaud, S., Blanchard, C., Martin, B., Guillot, P.V. & Duquenne, V. (1998) Neuronal and behavioral differences between Mus musculus domesticus (C57BL/6JBy) and Mus musculus castaneus (CAST/Ei). Behav Brain Res 95, 135–142.

Manser, C.E., Morris, T.H. & Broom, D.M. (1994) An investigation into the effects of solid or grid cage flooring on the welfare of laboratory rats. Laboratory Animals 29, 353–363.
Paigen, K. & Eppig, J.T. (2000) A mouse phenome project. Mammalian Genome 11, 715–717.
Richter, C.P. (1954) The effects of domestication and selection on the behavior of the Norway rat. J Natl Cancer Inst 15, 727–738.
Rustay, N.R., Wahlsten, D. & Crabbe, J.C. (in press) Influence of task parameters on rotarod performance and sensitivity to ethanol in mice. Behav Brain Res.
Smith, R.H. & Connor, J.L. (1974) The inheritance of behavioral wildness in house mice (Mus musculus L.). Anim Learning Behav 2, 249–258.
Stone, C.P. (1932) Wildness and savageness in rats of different strains. In Lashley, K.S. (ed.), Studies in the Dynamics of Behavior. University of Chicago Press, Chicago, pp. 1–55.
Wahlsten, D. (1973) Contributions of the genes albinism (c) and retinal degeneration (rd) to a strain-by–training procedures interaction in avoidance learning. Behav Genet 3, 303–316.
Wahlsten, D., Metten, P., Phillips, T.J., Boehm, S.L.I.I., Burkhart-Kasch, S., Dorow, J., Doerksen, S., Downing, C., Fogarty, J., Hen, R., McKinnon, C.S., Merrill, C.M., Nolte, C., Schalomon, P.M., Schlumbohm, J.R., Sibert, J.R., Wenger, C.D., Dudek, B.C. & Crabbe, J.C. (2003) Different data from different labs: lessons from studies of gene–environment interaction. J Neurobiol 54, 283–311.
Yerkes, R. M. (1913) The heredity of savageness and wildness in rats. J Anim Beha v 3, 286–296.
Acknowledgments
This work was supported by the Mouse Phenome Project, NIH grants R01 AA1 2714 and P50 AA1 0760, NSERC grant 45825 and the Department of Veterans Affairs. The authors are grateful to Courtney Bryden, Brandie Moisan, Sean F. Cooper, Jason Schlumbohm and Chia-Hua Yu for their adroit handling of the mice.